



Edmund Chattoe-Brown (2011)

Two Challenges in Simulating the Social Processes of Science

Journal of Artificial Societies and Social Simulation 14 (4) 1
<<http://jasss.soc.surrey.ac.uk/14/4/1.html>>

Received: 08-Jun-2011 Accepted: 19-Jul-2011 Published: 31-Oct-2011



Abstract

This note discusses two challenges to simulating the social process of science. The first is developing an adequately rich representation of the underlying Data Generation Process which scientific progress can "learn". The second is how to get effective data on what, in broad terms, the properties of the "future" are. Paradoxically, with due care, we may learn a lot about the future by studying the past.

Keywords:

Simulating Science, Algorithmic Chemistry, Evolutionary Algorithms, Data Structures, Learning Systems



Introduction

- 1.1 We can view science (leaving aside non-empirical fields like Post Modernism and Social Theory) as a discovery process. There is something "out there" which we use a set of shared practices and procedures to build a representation of. This is easiest to see in fields like physics where a combination of measuring devices, experiments and theories allow us to say things about the cooling of a metal bar or the failure of a wire under tension. This view of science raises two linked challenges for simulation.



The First Challenge: How Can We Represent Science in a Simulation?

- 2.1 The first challenge is: What is an adequate representation of the "something out there" to challenge any model of the discovery process. It is no use having a very realistic model of the discovery process trying to learn a very unrealistic "environment" or the results are likely to be nonsense (and not even usable in principle for "policy"). In this sense, unrealistic means very structurally unlike the "raw material" of real science. In a conceptually novel paper, Birchenhall and Windrum (1998) modelled the process of scientific discovery as the learning of a *function*, with different scientists working on the function at different levels of detail, over different ranges of values, exchanging information and so on. This is a distinctive use of simulation. Because we know the Data Generation Process (which we usually cannot observe directly in social science) we can make substantive claims about the effectiveness of learning processes: "This learning system will tend to discover underlying functions up to *this* level of difficulty". This means that we are no longer obliged to say exactly what the form of the scientific "discovery" function is (which is lucky because it is still unfolding) but only what our reasons are for putting it into a certain difficulty class (which connects to the second challenge below). This approach has "real" implications too. It is well known that "evolutionary" learning processes can cope better with difficult search spaces than, for example, simple hill climbing. Thus certain features of the scientific community (multiple teams working on the same problem, some scientific "competition" to drive people to new approaches or variations on existing ones and so on) appear to be capable of giving the advantages of evolutionary search in the system taken as a whole.
- 2.2 Nonetheless, it is useful to think more carefully about why the business of science cannot necessarily be represented accurately as a simple function. In dealing with novelty, we must actually develop the concepts (and relations between concepts) we need to build the representation of what is "out there" pretty much from the ground up. (Something now as "every day" as temperature was itself, at some point, a concept that needed to be rendered scientific and operationalised for use as a basis in other scientific work. This is one of the challenges that social science faces. It lacks relatively uncontroversial "measurement building blocks" equivalent to length, mass and temperature in physics. This is because the regularities of physics also apply to its measuring devices while the same is not true in social science.) This suggests the role of some kind of "dynamic concept network" representation of scientific knowledge (as when an overarching theory based on a new representation may "subsume" a whole sequence of apparently disparate research based on older or disparate representations). Secondly, because science is

unavoidably bound up with novelty, we may need to use relatively neglected modelling methods like algorithmic chemistry (Fontana 1991) that can actually cope with the arrival of genuinely new elements in the system (and transformations that only become possible with the arrival of those elements). Thirdly, we cannot treat science as merely an individualised set of papers or concepts but have to acknowledge the role of institutions as potentially divergent communities of practice and socialising entities. Thus the first challenge is to select (or develop) a relatively simple representation of science on which learning systems intended to model the scientific process can be compared. This representation needs to be relatively tractable, but not at the expense of failing to reproduce the stylised features of scientific progress as we currently know them: Citation patterns, punctuated equilibria after major discoveries, long latencies for the "application" of at least some theoretical ideas and so on. Meeting this first challenge also gives rise to the second.



The Second Challenge: How Can We Get Data About The Future?

- 3.1 The second challenge is to develop an effective approach to providing data for a simulation of the scientific process. In particular, it is very important not to confuse an ability to predict "the future" (which is widely considered to be the goal of science) with an ability to predict "futures" relative to the time period over which a model has been calibrated. By this I mean that we have considerable amounts of historical data on which models can be developed and then "tested" even though the predictions made are actually about events that took place in what is, for us (now), the past. It doesn't seem completely unreasonable that there is some underlying regularity to patterns of discovery (for example a Zipf distribution on citation patterns identifying the most important discoveries that happen at more or less similar intervals). Clearly, we cannot say what the next major breakthrough will be, or exactly when it will occur but we might say, with some plausibility if we use past data, that a model good enough to predict existing patterns might also suggest ways in which "major" discoveries could be made more frequent or "lost" discoveries be more reliably avoided.
- 3.2 Thus the challenge of building a representation of unfolding knowledge which can map onto the "out there" sought by science can be supported by its ability to represent the processes of formalisation, reconceptualisation, measurement and experiment which can be observed in historical case studies of science. If we cannot represent what we know has already happened reasonably economically in our abstraction then we are unlikely to find it much use for dealing, even approximately, with what *might* happen. Thus, paradoxically, to build models capable of understanding the future, we should look (particularly in "old" domains like science) very hard at the past. (In this vein, but apparently apocryphally, the bank robber Willie Sutton is supposed to have said "because that's where the money is" when asked why he robbed banks!)
- 3.3 Comparative models of the scientific process which explored its learning potential on a realistic representation of scientific knowledge, developed inductively from historical data, would stand at least a reasonable chance of doing what might appear impossible, allowing us to "model" and compare the efficacy of policies for doing something which, by its nature, is continually unknowable in its details until it happens.



References

- BIRCHENHALL, C. R. and Windrum, P. (1998). Developing simulation models with policy relevance: getting to grips with UK science policy. In P. Ahrweiler & N. Gilbert (Eds.), *Computer Simulations in Science and Technology Studies* (pp. 183-206). Berlin: Springer-Verlag.
- FONTANA, W. (1991). Algorithmic chemistry. In C. G. Langton, C. Taylor, J. D. Farmer & S. Rassmussen (Eds.), *Artificial Life II: Proceedings of the Workshop on Artificial Life, Santa Fe, New Mexico, February 1990*, SFI Studies in the Sciences of Complexity, Proceedings Volume X (pp. 159-209). Redwood City, CA: Addison-Wesley.